O w]

SEQUENCE LISTING

GENERAL INFORMATION: (1)

(i) APPLICANT:

Lebrun, Michel Sailland, Alain Freyssinet, Georges DeGryse, Eric

(ii) TITLE OF INVENTION: Mutated 5-enolpyruvylshikimate-3-phosphate synthese, gene coding for this protein and transformed plants containing this gene

(iii) NUMBER OF SEQUENCES: 5

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Connolly, Bove, Lodge and Hutz

(B) STREET: 1220 Market Street

(C) CITY: Wilmington

(D) STATE: Delaware

(E) COUNTRY: U.S.A.

(F) ZIP: 19899

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE:

(B) COMPUTER:

(C) OPERATING SYSTEM:

(D) SOFTWARE:

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PCT/FR96/01128

(B) FILING DATE: 18-JUL-1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Robert G. McMorrow, Jr.

(viii) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (302) 658-9141

(B) TELEFAX: (302) 658-5614

(2) INFORMATION FOR SEQ ID NO: 1/2

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1713 Same pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: /inear

(11) HOLECULE TYPE: CONA

(vi) ORIGINAL BOURCE: (A) ORGANISM: See mayo (B) STRAIN: Black Mexican Sweet

(F) TIESUE TYPE: Callus

(vii) Decisiate sounce: (A) Library: lambda gt10

JE) CLONE: papa-HL-711

(xX) SEQUENCE DESCRIPTION: SEQ ID MO:1:

,						
MICATTIC	KCKCMONN	CASCTATOAC	CATGATTACS	AATTOGGGCC	COGOCOCOTO	60
ATOOSSOSSIC	GCAGCEGCS	6000000790	A0000007700	COMOGRAÇATO	GTGCTGCAGC	120
CATCAACCA	CATCTCCGGC	ACCOTTCAMAC	T00000007C	CAMPIOSCIT	TOCALODISA	180
דסגדססדואנד	0000000000000000	TOOSAGGGGA	CALCUSTOST	TEATAACCTG	CTGAACAGTS	240
AGGATGTCEA	CTACATGCTC	GGGGCCTTGA	GGACTCTTGG	TCTCTCTOTC	GAAGCOGACA	300
MOCTOCCAA	AAGAGCTGTA	аттаттааст	GTGGTGGAAA	onto contra	GAGGATOCTA	360
NGAGGAAGT	OCADCTCTTC	77000GAA70	CTGGNACTGC	AATOCOGCCA	TTGACAGCAG	420
	70770770CA			**********	AGA MICA COC	480

#12

MIANACCEAE TRIBENACTER OPERITORIAT TGANICARIT TONTOCAGAS OTTGATEGET	54
TESTIFICAC TRASTRISCA CONTINUON TEAATRIAAT CEGALIGOTA CONGETECA	•10
AUDITOMAGET ITETITIVETOS ATGANICATIO ANTACTINAG TROCTINETIS ATRICCISETE	64
CITTINGCTOT TOURGATHING GAGANITHAAA TCATTGATAA ATMANICTOU ATTCCGTACG	72
TOGAMATIGAC AITTGAGAITG ATGGAGOTT TTGGTGTGAA AGCAGAGCAI TCTGATAGCT	78
COGREMANT CTACATTRAG GGAGGTCANA ANTACRAGTC COCTAMART GOCTATOTTG	840
AAGGTGATGC CTCAAGCGCA AGCTATTTCT TGGCTGGTGC TGCAATTAGT GGAGGGACTG	900
TEACTOTOGO ACCACCACTT TECAGOSTOA TETGAACTTT GCTGAGGTAC	960
TIGGAGATGAT GOGAGCIGAAG GTTACATIGGA COGAGACTAG COTAACTIGTT ACTIGGCOCCAC	1020
COCEGGRAGOC ATTEGGRAGE ANACACCECA AGECGATEGA TOTCAACATE ANGAAGATEC	1080
CTGATGTOGC CATGACTCTT GCTGTGGTTG COCTCTTTGC CGATGGCCCG ACAGCCATCA	1140
GAGACITIGGC TTCCTGGAGA GTAAAGGAGA CCGAGAGGAT GGTTGCGATC CHGACGGAGC	1200
TANCCARGET GOBAGCATCT GTTGAGGAAG GGCCOGACTA CTGCATCATC ACCCGGCGG	1260
AGAAGGTGAA COTGACIGCG ATCGACACGT AGGACGAGCA CAGGATGGCC ATGGCCTTCT	1320
COUTTIGUESC CTIGTECCEAS OTCCCCCTICA CCATOCEGAA COUTGGOTGC ACCEGAAGA	1380
extrecessa ctactiosal stocisasca cittosteaa gaatlaataa asostosaa	1440
ACTACCACEC AGCTTGATTG AAGTGATAEG CTTGTGCTGA GGAAATACAT TTCTTTTGTT	1500
PIGITITICI CITICACOGG ATIANGITTI GAGICIGIAA COTTAGITGI TIGIAGCAAG	1560
TICIATITC GGATCITANG TITGIGCACT GIANGCCANA TITCATITCA AGASTGGTTC	1620
TTCGAATAA TAAGAATAAT AAATTACOTT TCASTGAAAA AAAAAAAA AAAAAAAA	1680
MANAMA AAAAAAAA AACCCGGGAA TTC	1713
21 TURNOMARUNI DOR AND DO	2123

- (2) INFORMATION FOR SEQ ID NO.2:
 - (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1340 base pairs
 - (B) TYPE: nucleic scid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
 - (11) MOLECULE TYPE: CONA

 - (V1) ORIGINAL SOURCE:
 - (A) ORGANISM: Zes Bays (B) STRAIN: Black Mexican Sweet
 - (VIII DOCULATE SOURCE:
 - (B) CLONE: PRPA-NL-716
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 6..1337
 - (x1) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCATG	GCC Ala 1	GGC Gly	GCC Ala	GAG Glu	GAG Glu 5	ATC Ile	G7G Val	CTG Leu	CAG Gln	ccc ₽ro 10	ATC Ile	AAG Lys	GJ <i>n</i> GVG	ATC 11e	;	47
700 G	3C A	x 0	C A	va (=	× ~	72 00	-	~								

TOO GGC ACC GTC AAG CTG COG GGG TCC AAG TCG CTT TCC AAC CGG ATC Ser Gly Thr Vol Lym Leu Pro Gly Ser Lym Ser Leu Ser Aen Arg Ile 15 20 25 30

CTC CTA CTC GCC GCC CTG TCC GAG GGG AGA ACA GTG GTT GAT AAC CTG Leu Leu Leu Ala Ala Leu Ser Glu Gly Thr Thr Val Val Asp Asn Leu 143

CTG AAC AGT GAG GAT GTC CAC TAC ATG CTC GGG GCC TTG AGG ACT CTT Lou Agn Ser Glu Asp Val His Tyr Het Leu Gly Ale Lou Arg Thr Leu 191

50		55	60	
GTT CTC TCT (TCC GLY Law Sec Val 69	CIN VIS VAD	AAA (ICT (ICC A Lys Als Als L 70	AA AGA GCT GTA GTT G YS AFG Ala Val Val V 75	ति 239 'वर
80	85	AST OTH WED W	T AAA GAG GAA GTG C S inv uit uit Gud AAA TG 90	ln
95	100	INF ALE MOE A	-	1 a 10
	115	120	G CTT GAT GGA GTA CX 1 Leu Asp Gly Val Pi 125	10
130	wid had 114 (135	T GTC GGA TIG ANG CF 1 Vel Gly Leu Lys Gl 140	.n
145	TEL MED CYS I	ne Leu Gly Th: 150	T GAC 76C CCA CCT GT F Asp Cyw Pro Pro Ve 185	1
160	165	en sue età GT	ANG GTC ANG CTG TC / Lys Val Lys Lau Se 170	•
175	180	782 An set VIE For	731	5
1	.95	200	GAT AAA TTA ATC TCC Amp Lym Leu Ile Ser 205	•
210	TO HEE THE D	215	GAG COT TIT GOT GTG Glu Ary Phe Gly Vel 220	•
225	21 21	D Veb Vtd bye	TAC ATT ANG GGA GGT Tyr Ile Lye Gly Gly 235	,
240	245	n Als Tyr Val	GAA GOT GAT GOC TCA Glu Gly Asp Als Ber 250	
255	260	265	ACT COA GGG ACT CTC Thr Gly Gly Thr Val 270	815
27	5	P BOP Lau Gin 280	GGT GAT GTG AAG TTT Gly Amp Val Lym Phe 285	863
290	G MAT WAS GT	295	ACA TOO ACC OAS ACT The Trp The Glu The 300	911
305	110 110 110 110	Arg Glu Pro	TTT GGG AGG AAA CAC Phe Gly Arg Lye His 315	989
320	3,24	Van The Mer I	180	1007
ACT CIT GCT GTG GTT Thr Leu Ala Val Val 335 GAC GTG GCT TCC TGG	340 Den Iue	342	o Thr Ala Ile Arg 350	1055
GAC GTG GCT TCC TGG Asp Val Ala Ser Trp 355	MA AST TAR	360	rg Met Vel Ala Ila 365	1103
CGG ACE GAS CTA ACC Arg Thr Glu Leu thr	The ran City	WENT TOT OFF G	THE GIV GIV ELS PED	1151

170 THE THE ATE ATE ACT OF ON THE AGE CTO ACT OF ATE THE Tye Cyn Ilu Ilu Thr Pes Pro Olu Lys Leu Am Val Thr Ala Ile Asp 1199 ACT THE CHAC CAS AND ATT GOD ATT GOD THE THE TYP AMP AMP HER ACT HER ALS NOT ALS PHO SUE LEG ALS ALS CYS 400 1.17 GCC GAG GTC CCC GTC ACC ATC CGG GAC CCT GGG TGC ACC CGG AAG ACC Ala Glu Vel Pro Vel Thr Ile Arg Asp Pro Gly Cye Thr Arg Lye Thr 415 420 425 430 1295 TTC CCC GAC TAC TTC GAT GTG CTG AGC ACT TTC GTC AAG AAT Phe Pro Asp Tyr Phe Asp Vel Leu Ser Thr Phe Vel Lys Asn 1337 440

(2) INFORMATION FOR SEG ID NO:3:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 444 sains scids
- (B) TYPE: amino maid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Ale Gly Ale Glu Glu Ile Vel Leu Gln Pro Ile Lye Glu Ile Ser Gly
1 5 10 15

Thr Val Lys Lau Pro Gly Ser Lys Ser Lau Ser Asn Arg Ile Lau Lau 20 25 30

Leu Als Als Leu Ser Glu Gly Thr Thr Val Val Asp Asn Leu Leu Asn 35 40 45

Ser Glu Asp Val His Tyr Met Leu Gly Ala Leu Arg Thr Leu Gly Leu 50 55 60

Ser Val Glu Ala Aap Lys Ala Ala Lys Arg Ala Val Val Gly Cys 65 75 80

Gly Gly Lys Phe Pro Val Glu Asp Ala Lys Glu Glu Val Gln Leu Phe 85 90 93

Leu Gly Asn Ala Gly Thr Ala Met Arg Pro Leu Thr Ala Ala Val Thr

Ala Ala Oly Gly Asn Ala Thr Tyr Val Leu Asp Gly Val Pro Arg Met 115 120 125

Arg Glu Arg Pro Ile Gly Asp Leu Val Vel Gly Leu Lys Gln Leu Gly 130 135 140

Alm Amp Val Amp Cym Phm Leu Gly Thr Amp Cym Pro Pro Val Arg Val 145 150 150 160

Asn Gly Ile Gly Gly Leu Pro Gly Gly Lys Val Lys Leu Ser Gly Ser 165 170 175

Ile Ser Ser Gln Tyr Lou Ser Ala Lau Leu Met Ala Ala Pro Leu Ala 180 185 190 Leu Gly Asp Val Glu Ile Glu Ile Ile Asp Lys Leu Ile Ser Ile Pro 195 205

Tyr Val Glu Het Thr Leu Arg Leu Het Glu Arg Phe Gly Val Lye Ala 210 215 220

Glu His Ser Asp Ser Trp Asp Arg Phe Tyr Ile Lye Gly Gly Gln Lys 225 235 240

Tyr Lym Ser Pro Lym Amn Ala Tyr Val Glu Gly Amp Ala Ser Ser Ala 245 250 255

4.	r T	YE	[Pro-	1414 240	i ALı	i GLY	/ A14	V14	245	Thr	. UIA	uta	The	. Val		r Val	
GI	u G	lγ	Cys: 375	σιγ	'the	The	S96	230 201	i Gin	GLY	/ Asp	Val	Lya Lya		Ala	Glu	
	•	14.1					295	1			'f rp	300					
30	•					310					315					120°	
					325					330					335		
				340					345		Thr			350	_		
			100					360			Met		365		_		-
	3,	U					375				Glu	380					
38:	i II	• 7	he	\$100	510	G1u 390	Lys	Lau	Asn	Va1	Thr 395	Ala	Ile	Asp	Thr	Tyr 400	
λας	As	p H	is	Arg	Mat 405	Ale	Met	Ale	Phe	8es 410	Leu	Al _e	Ale	Cys	Als 415	Glu	
Val	. Pr	o V	a l	7hr 420	Ile.	Arg	Asp	Pro	Gly 425	Cys	Thr	γιΑ	Lys	7hr 430	Phe	Pro	
Asp	7у.	7 P	he 35	Asp	Val	Leu	Ser	Thr 440	Phe	Va1	Lyp	Asn					
(2)	IN	POR	MAT	ION	FOR	6EQ	ID N	: 4 :									
	(:	1)	IA IB) LE) TY ST	noth Pe :	i: 13 Duel Bone	40 b eic 55:	acid doub	peir	•							·
					E TY												
	(٧)	L) ((A)) QR	l 80 Dani Rain	BH:	les :		cen :	Byrae	t						
	(*11	.) 1			te 8 One:			-720									
	(ix	i) [A		: Œ/KI ZA7I(337									
	(×i) 6	EQL	ÆKZ	DE:	BCR I	PZIO	N: 81	II QI	NO:	4:						
CCAT	o on	oc la l	GC C	ALE ALE	GL	61/ 61/	ATC J Ile	GT(G CTG	G)	Pro	Ile	Lyt	GAG Glu	ATC Ile	:	41
TCC Ser 15	gly	AC Th	r y	TC A	VAG C	779 (Jeu 6 20	203 (Pro (igg 1	er i	ya s	103 C ler 1 25	TT 1 Ou 8	er A	uac c Ven a	rg J	110 30	9;
CTC Leu	CTA Lou	CT La	C G	78 W	00 c	770 T	er d	AG G	Hy T	CA # hr 1 40	CA G	7G G al V	TT G	usp A	AC C an L 45	77G ÆU	141
CTG Leu	AAC Aan	AG Se	r G	AG (3 1u A 50	AT G ap V	TC C	ac t	AL H	arg C let L 55	TC q	as s ly A	CC T	ou A	GG A Ly 7 60	or c	 	191
GOT Gly	CTC Leu	TC 80.	r v	TC G	AA G lu A	ya Y CC C	ap L	AA G	CT G	œ A	AA A Yo A	4 65	ता ज 1: V	TA G	IT G	77 61	239

										GCT Ala						287
										СіХі Ага 105						335
										ord Val					CCA Pro-	383
AGA Arg	ATG Het	agg Afg	6A6 61u 130	aga Arg	Pro Pro	ATT 11e	01 y GGC	GAC Asp 135	TTG Leu	011 V=1	GTC Val	GCIA CJ1 y	TTG Leu 140	AAG Lys	Gln Gln	431
CTT Leu	GGT Gly	GCA Ala 145	GAT Asp	GTI Val	GAT Asp	CAR 101	TTC Phe 150	CTT Leu	GGC Gly	ACT The	ØAC Asp	TGC Cys 155	CCA Pro	CCT Pro	017 Val	479
										GGC Gly						527
										776 Leu 185						575
TTG Leu	gct Ala	CTT Leu	GGG Gly	GAT Asp 195	gig Val	G A G Glu	ATT Ile	Glu	ATC Ile 200	ATT Ile	gat Asp	aaa Lys	TTA Leu	ATC 110 205	TCC Ser	623
										ATG Net						671
Lyb	GCA Ala	61u 225	CAT His	TCT Ser	GAT A≇p	AGC Ser	TGG T£p 230	gac Asp	Aga	TTC Phe	tac tyr	ATT 110 235	Lys AAG	gga Gly	gly gg7	719
										GTT Val						767
AGC Ser 255	GCA Ala	AGC Ser	TAT Tyr	TTC Phe	77G Leu 260	GCT Als	GGT Gly	GCT Ala	gca al 1	ATT 11e 265	ACT The	gly gga	01 y GGG	ACT Thr	GTG Val 270	815
ACT The	GTG Ve 1	GAA Glu	GOT Gly	TGT Cyn 275	GGC Gly	ACC Thr	ACC Th <i>s</i>	76A 108	77G Leu 280	CAG Gln	Gly	gat Asp	G7G Val	AAG Lys 285	TTT Pho	863
OCT Als	G A G Glu	GTA Val	CTG Leu 290	GAG Glu	ATG Met	ATG Het	GGA Gly	GCG Al = 295	AAG Lys	GTT Val	ACA Thr	TGG Trp	ACC 1hr 300	GAG Glu	ACT Thr	911
AGC Ser	OTA Val	ACT Thr 305	GTI Val	ACT Thr	gjy GGC	CCA Pro	Pro 310	Arg C3G	Glu Glu	019 019	TTT Phe	GGG Gly 315	AGG AEG	EK7 YYY	KT2 CVC	959
CTC	AAG Lys 320	GCS Als	ATT Ile	GAT Asp	GIC Val	AAC Asn 325	ATG Met	AAC Aan	aag Lys	ATG Not	CCT Pro 330	GAT Asp	GTC Vel	GCC Alm	ATG Met	1007
ACT Thr 335	CTT	AL:	OTG Val	GTT Val	000 Ala 340	r=a CIC	TTT Phe	dCC Ale	GA? Asp	GGC Gly 345	003 Pro	ACA The	A7# GCC	ATC 11s	AGA Arg 350	1055
GAĆ Asp	oro Val	GCT Ala	TCC Ser	766 7rp 355	ADA	OTA Val	AAG Lys	G)u	7hr 360	Glu Glu	YEG YGG	ATG Met	GTT Val	906 Ala 365	ATC Ile	1103
Yt.d COG	ACG Thr	GAG Glu	CTA Lou 370	ACC The	AAG Lys	C7G Leu	gga Gly	6CA Ale 375	er Ser	GTT Vel	GNG Glu	GAA Glu	606 617 380	000 P.FO	GAC Asp	1151
TAC Tyr	TGC Cys	ATC 11e 385	Ile	ACG Thr	CCG Pro	CC3 Pro	GNG Glu 390	AAG Lys	ren C10	AAC Asn	GTG Val	ACG Thr 395	V) •	ATC Ile	GAC Amp	1199

ACT TAC TAC TAC TAC CAC AND ATT GCT ATT GCC TTC TCC CCT CCC GCC TGT
The Tyr Aip Aip Hix Ard Net Ala Net Ala Phe Ser Leu Ala Ala Cys
400

GCCT MAG GTC CCC GTC ACC ATC GCC GAC GCC TGT GCC TGC GCC TGT
ALA GTU MAI Pro Mai The He Ard Aip Pro Gly Cys The Ard Lya The
420

TTC CCC GAC TAC TTC GAT GTG CTG AGC ACT TTC GTC AAG AAT
Phe Pro Asp Tyr Phe Asp Mai Leu Ser The Phe Mai Lya Ash
435

(2) INFORMATION FOR SEQ ID NO:5:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 444 amino society
 - (8) TYPE: smino scid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:5:
- Als Gly Als Glu Glu Ile Vel Leu Gln Pro Ile Lye Glu Ile Ser Gly
 1 5 10 15
- Thr Val Lys Leu Pro Gly Ser Lys Ser Leu Ser Asn Arg Ile Leu Leu 20 25 30
- Leu Ala Ala Leu Ser Glu Gly Thr Thr Val Val Asp Asn Leu Leu Asn 35 40 45
- Ser dlu Asp Val His Tyr Met Leu Gly Ala Leu Arg Thr Leu Gly Lau 50 55 60
- Ser Vel Glu Ala Asp Lys Ala Als Lys Arg Ala Vel Vel Vel Gly Cys 65 75 80
- Gly Gly Lys Phe Pro Val Glu Asp Ala Lys Glu Glu Val Gln Leu Phe 85 90 95
- Leu Gly Asn Ala Gly Ile Ala Net Arg Ser Leu 7hr Ala Ala Val Thr 100 105 110
- Ala Ala Gly Gly Amn Ala Thr Tyr Val Leu Amp Gly Val Pro Arg Met 115 120 125
- Arg Glu Arg Pro Ile Gly Asp Leu Val Val Gly Leu Lys Gln Leu Gly 130 140
- Ala Asp Vel Asp Cys Phe Leu Gly Thr Asp Cys Pro Pro Vel Arg Vel 145 150 155 160
- Asn Gly Ile Gly Gly Lau Pro Gly Gly Lys Val Lys Leu Ser Gly Ser 165 170 175
- The Ser Ser Gin Tyr Leu Ser Als Leu Leu Met Ala Ala Pro Leu Ala 180 185 190
- Leu Gly Asp Val Glu Ile Glu Ile Ile Asp Lys Leu Ile Ser Ile Pro 195 200 205
- Tyr Vel Glu Net Thr Lau Arg Lau Net Glu Arg Phe Gly Val Lys Ale 210 225 220
- Glu His Ser Asp Ser Trp Asp Arg Phe Tyr Ile Lys Gly Gly Gln Lye 225 235 240
- Tyr Lys Sar Pro Lys Ash Ala Tyr Vel Glu Gly Asp Ala Ser Ser Ala 245 250 255
- Ser Tyr Phe Leu Ale Gly Ala Ale Ile Thr Gly Gly Thr Vel Thr Vel 265 270
- Glu Gly Cya Gly Thr Thr Ser Lau Gln Gly Asp Val Lys Phe Ala Glu 275 280 280

Val 240 Glu Met Mer Gly Ala Lyx Val Thr Trp Thr Glu Thr Ser Val 300 Glu Thr Ser Val 315 Val Thr Gly Pro Pro Ard Glu Pro Pho Gly Ard Lyx His Lew Lyx Jab Gle Axp Val Axp Met Axp Lyx Met Axp Val Axp Met Axp Jab Axp Val Axi Met Thr Lew Jab Axp Val Axi Val Axi Met Thr Lew Jab Axp Gly Pro Thr Axi Ite Ard Axp Val Axi Met Thr Lew Jab Axp Val Axi Met Thr Lew Jab Axp Val Axi Met Thr Lew Jab Axi Val Axi Met Axp Val Axi Met Axp Val Axi Met Axp Val Axi Met Thr Lew Jab Axp Val Axi Met Axp Xxi Met Val Axi Ite Ard Axp Val Axi Met Axi Met Axp Thr Axi Met Axi Me

٠.,